

TECHNICAL REPORT
Multivariate F-factor Sampling Based on Objective Bayesian Statistics
REX study
Lack of Sufficiency=0.5, Necessity=0.5

September 6, 2016

Contents

1 Problem Statement **3**

2 Objective Bayesian Statistics **4**
2.1 Reference prior . . . 4
2.2 The case of normal location parameters . . . 4

3 Example of a link function **5**

4 Dataset Summary **6**
4.1 Sample size and basic statistics . . . 6
4.2 Biomarker Boxplots . . . 7
4.3 Checking the Betweenness Assumption . . . 8

5 Posterior distribution for RD_i **9**

6 Sampling the link functions **10**
6.1 Sampling the β's . . . 10
6.2 Sampling the main coefficients . . . 10

7 Results **10**
7.1 Final distribution for F . . . 10
7.2 Example of application to the PHIM . . . 11

8 Annexes **12**

List of Figures

1 Example of a link function . . . 5
2 Boxplots the endpoints under consideration . . . 8
3 Posterior distributions for the endpoints under consideration . . . 9
4 Overall Posterior distribution from the sampled link functions . . . 11
5 ER model for F sampled from the density estimate . . . 12

List of Tables

1 Sample size for each arm and each biomarker. . . . 6
2 Descriptive statistics for the biomarkers in each arm. Med: median, Q1: first quartile, Q3: third quartile. . . . 7
3 Individual sufficiency values. . . . 10

1 Problem Statement

The F-factor is defined as the parameter in the exponential negative population model for the excess risk (ER) for smoking cessation and switching to RRP.

$$ER_{Switch}(a,t) = ER_{CC}(a)(F + (1-F)e^{-\frac{\log(2)\cdot t}{H}})$$

where t is the time from switching, H is the disease specific half life and a is the age. We have:

$$ER_{Switch}(a,t) - ER_{Cess}(a,t) = ER_{CC}(a,t) \cdot F \cdot [1 - e^{-\frac{\log(2)\cdot t}{H}}]$$

and

$$ER_{CC}(a,t) - ER_{Cess}(a,t) = ER_{CC}(a,t) \cdot [1 - e^{-\frac{\log(2)\cdot t}{H}}]$$

Therefore:

$$\frac{ER_{Switch}(a,t) - ER_{Cess}(a,t)}{ER_{CC}(a,t) - ER_{Cess}(a,t)} = F$$

For a given endpoint (biomarker) X, the difference in biomarker concentration for $G = CC$ ($Cess$ or $Switch$ respectively) is defined as $\mathbb{E}(X|G) - \mathbb{E}(X|NS)$, where NS denotes the never smoker group. The relative biomarker change is therefore

$$\frac{(\mathbb{E}(X|Switch) - \mathbb{E}(X|NS)) - (\mathbb{E}(X|Cess) - \mathbb{E}(X|NS))}{(\mathbb{E}(X|CC) - \mathbb{E}(X|NS)) - (\mathbb{E}(X|Cess) - \mathbb{E}(X|NS))} = \frac{\mathbb{E}(X|Switch) - \mathbb{E}(X|Cess)}{\mathbb{E}(X|CC) - \mathbb{E}(X|Cess)}$$

Consequently, if X_i is any endpoint measured in the context of a switching design, the relative decrease is defined as

$$RD_i \doteq \frac{\mathbb{E}(X_i|Switch) - \mathbb{E}(X_i|Cess)}{\mathbb{E}(X_i|CC) - \mathbb{E}(X_i|Cess)}$$

As the study design is longitudinal, this ratio is equivalent to

$$RD_i = 1 - \frac{\mathbb{E}(\Delta(X_i|Switch))}{\mathbb{E}(\Delta(X_i|Cess))}$$

where $\Delta(X_i)$ is the baseline corrected value $X_i(Day_n) - X_i(Day_0)$; under the assumption that $\mathbb{E}(\Delta(X_i|CC)) = 0$ (i.e., no change in average for the biomarkers between Day n and Day 0) . In absence of further information our belief about F is $\mathcal{U}(0,1)$ (the "uninformed" F). To update this belief, we expect some of information about the parameter F to be contained in RD_i. A list of assumptions made is listed below:

1. *Normality* The conditional distribution of X_i within the group G (G being Switch, CC or Cess) is $X_i|G \sim \mathcal{N}(\mu_i(G), \sigma_i(G))$;
2. *Betweenness* One of the orderings below holds:
   (a) $\mu_i(CC) \geq \mu_i(Switch) \geq \mu_i(Cess)$ and $\mu_i(CC) \geq \mu_i(Cess) + \xi_i$,where $\xi_i > 0$;
   (b) $\mu_i(CC) \leq \mu_i(Switch) \leq \mu_i(Cess)$ and $\mu_i(CC) \leq \mu_i(Cess) - \xi_i$,where $\xi_i > 0$. In this case, $-X_i$ will be considered instead.
3. *Informativeness* $(RD_1, ..., RD_n)$ is informative on F in the form of $F = \psi(RD_1, ..., RD_n) + \alpha_0\psi_0(U)$ with ψ and ψ_0 measurable. $\psi_0(U)$ is the link function of an unknown mutli-endpoint biomarker that would ensure the sufficiency of the total set $(RD_1, ..., RD_n, U)$. **The lack of sufficiency of the multi-endpoint biomarker $(RD_1, ..., RD_n)$ to be a disease *risk* biomarker is therefore characterized by $\alpha_0 \cdot \psi_0(U)$.** The set of markers is considered to be necessary. Additionally, we have:
   (a) $\psi(1, ..., 1) = 1$
   (b) $\psi_0(U) \sim \mathcal{U}([0,1])$ follows an uniform distribution between 0 and 1.
   (c) *Conservativeness*: $\psi(RD_1, .., RD_n)$ can not underestimate F; i.e., $\psi_i(x_1, .., x_n) \geq (x_1, .., x_n)$ coordinate-wise; this translates the idea that no RD_i can be fully predictive of F and the future excess risk decrease.

(d) ψ is factorizable i.e., $\psi(RD_1, ..., RD_n) = \sum_i \alpha_i \psi_i(RD_i)$ with

i. *Monotonic coherence* A decrease in RD_i can not be indicative of an increase of F (the lower the relative decrease, the lower $\psi_i(RD_i)$: $\psi_i(x) \geq \psi_i(y)$ if $x \geq y$ defined on $[0,1]$ with image in $[0,1]$;

ii. $\psi_i(1) = 1$;

iii. $\alpha_i \geq 0$ $(i = 0, ...n)$ and, as $\psi(1, ..., 1) = 1$ we must have $\sum_{i=0}^{n} \alpha_i = 1$. **The α's describes the relative weight of each biomarker**; therefore α_0 quantifies the importance of the missing set of biomarker to reach sufficiency.

(e) The base functions ψ_i in the expansion of ψ are expressed as linear splines over $[0,1]$ of the form:

$$\psi_i(\beta_i, x_i) = \begin{cases} \frac{1}{\beta_i} x & \text{if } x_i < \beta_i \\ 1 & \text{otherwise} \end{cases}$$

The parameter β_i describes the minimum RD_i needed to draw a conclusion on F strictly less than 1: if $RD_i < \beta_1^i$ then $\psi_i(RD_i) < 1$ (i.e., quantifies the **necessity** of that biomarker i alone).

4. The link function does not depend on the age and disease;

As we can update our belief on the distribution of F using estimates of the probability distributions of RD (for a given link function), scenarios can be drawn from various choices of β_i and α_i. Therefore, the purpose is to draw samples from various scenarios for ψ based on objective Bayesian statistics estimation of the posterior distribution for RD_i. This framework is consistent with the fact that in absence of any information our belief about F is $\mathcal{U}(0,1)$. Indeed, if $\alpha_0 = 1$ (and hence $\alpha_i = 0, i = 1, .., n$; meaning the set of endpoints considered is not informing on F), then the final distribution of F will be $\mathcal{U}(0,1)$.

2 Objective Bayesian Statistics

2.1 Reference prior

We need to identify a non-informative prior mathematically. The *reference prior* is defined as the (possibly improper) prior that maximizes the missing information about the quantity of interest (i.e., the information which infinite independent replications could provide). More precisely, given a model $\{\mathbb{P}(\boldsymbol{x}, \theta)\}$, the amount of information which may be expected from $\boldsymbol{x}$ about θ is

$$I_\theta(\boldsymbol{X}, \pi(\theta)) = \mathbb{E}_{\boldsymbol{X}} \int_\Theta \pi(\theta|\boldsymbol{x}) \log \frac{\pi(\theta|\boldsymbol{x})}{\pi(\theta)}$$

Under some regularity constraints, the reference prior for a consistent asymptotically sufficient estimator of θ (univariate), π^*, is given by the so called Jeffrey's prior:

$$-\mathbb{E}(\frac{\partial^2 \log \mathbb{P}(\boldsymbol{x}|\theta)}{\partial \theta^2})^{\frac{1}{2}}$$

For the two parameters problem, θ, λ (e.g. μ, σ), we sequentially derive reference priors: first derive $\pi^*(\lambda|\theta)$. If the latter is proper, then $\pi^*(\theta)$ is obtained by integrating λ out $\mathbb{P}(x|\theta) = \int_\Lambda \mathbb{P}(\boldsymbol{x}|\theta, \lambda)\pi^*(\lambda|\theta)d\lambda$.

2.2 The case of normal location parameters

For $\mathcal{N}(\mu, \sigma)$, this leads to $\pi^*(\sigma|\mu) \propto \sigma^{-1}$ and $\pi^*(\mu) \propto 1$ and therefore $\pi^*_\mu(\mu, \sigma) = \sigma^{-1}$. Consequently, if the data $\boldsymbol{x}$ are from $\mathcal{N}(\mu, \sigma)$, the corresponding likelihood function is proportional to $\sigma^{-n} \cdot e^{-n\frac{s^2+(\bar{x}-\mu)^2}{2\sigma^2}}$, with $ns^2 = \sum(x_i - \bar{x})^2$ Using Bayes' theorem with the reference prior discussed above, the corresponding joint posterior is

$$\mathbb{P}(\mu, \sigma|\boldsymbol{X}) \propto \sigma^{-(n+1)} e^{-n\frac{s^2+(\bar{x}-\mu)^2}{2\sigma^2}}$$

Therefore using the Gamma integral in terms of $\lambda = \sigma^{-2}$, $\mathbb{P}(\mu|\boldsymbol{X})$ is recognized to be $St(\mu|\bar{x}, s/\sqrt{n-1}, n-1)$ (and $\mathbb{P}(\sigma|\boldsymbol{X}) \sim 2\sigma^{-3} Ga(\sigma^{-2}|(n-1)/2, ns^2/2)$). In our setting we choose the reference prior for each of the independent group (that can be considered as the reference prior for the coefficients of a linear model explaining the endpoints by the groups), also called *independent Jeffrey's prior*.

3 Example of a link function

As explicitly stated in our assumptions, the link will be parametrized as

$$\psi_{\alpha,\beta}(x) = \alpha_0\psi_0(u) + \sum_{i=1}^{n} \alpha_i\left(\frac{1}{\beta_i}\mathbb{I}_{[0,\beta_i]}(x_i) + \mathbb{I}_{[\beta_i,1]}(x_i)\right)$$

where $\mathbb{I}_A$ is the indicator function of the interval A. An example of a link function for a 2-endpoint biomarker, with $\alpha_0\psi_0(u) = 0.3$, $\alpha_1 = 0.5$ and $\alpha_2 = 0.2$ and $\beta_1 = 0.2$, $\beta_2 = 0.4$ is shown in Figure 1.

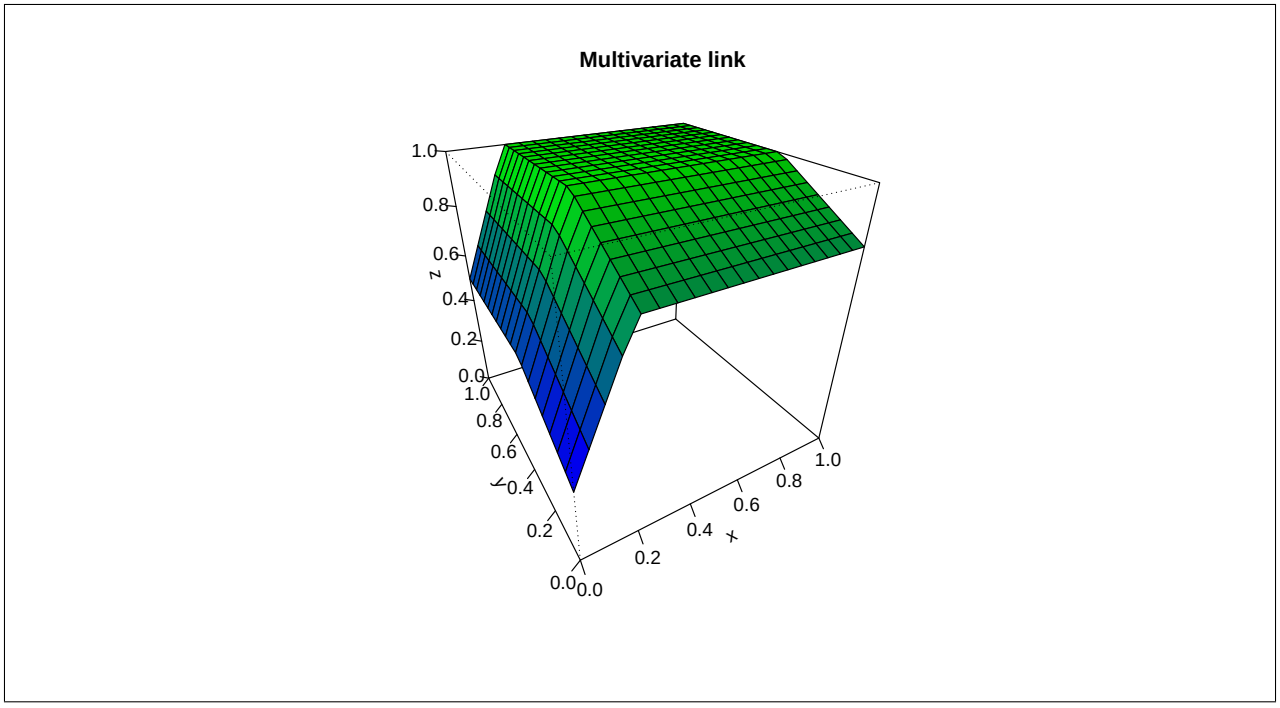


Figure 1: Example of a link function ψ; for $\alpha_0\psi_0(u) = 0.3$, $\alpha_1 = 0.5$ and $\alpha_2 = 0.2$ and $\beta_1 = 0.2$, $\beta_2 = 0.4$.

4 Dataset Summary

4.1 Sample size and basic statistics

The number of samples per arm considered in the study REX are shown in Table 1.

Table 1: Sample size for each arm and each biomarker.

	CC/mCC	THS/THSm 2.2	SA
(Day 5) CARBXHGB	157	309	139
(Day 5) CO	158	310	140
(Day 5) U1NACRE	157	308	138
(Day 5) U1OHPCRE	156	306	138
(Day 5) U2NACRE	157	308	138
(Day 5) U3HPMCRE	157	308	138
(Day 5) U4ABPCRE	157	308	138
(Day 5) UBAPCRE	157	308	138
(Day 5) UCEMACRE	157	308	138
(Day 5) UHEMACRE	157	308	138
(Day 5) UHMPMCRE	157	307	138
(Day 5) UMHBMCRE	156	306	138
(Day 5) UNNALCRE	157	308	138
(Day 5) UNNNCRE	157	307	138
(Day 5) UOTOLCRE	155	308	138
(Day 5) USPMACRE	156	306	138
(Day 90) 11-Dehydro-Thromboxane B2 (pg/mg creat)	73	117	46
(Day 90) 1-NA (pg/mg creat)	73	117	46
(Day 90) 2-NA (pg/mg creat)	73	117	46
(Day 90) 3-HPMA (ng/mg creat)	73	117	46
(Day 90) 4-ABP (pg/mg creat)	73	117	46
(Day 90) B[a]P (fg/mg creat)	73	117	46
(Day 90) CEMA (ng/mg creat)	73	117	46
(Day 90) Cholesterol (mg/dL)	73	117	46
(Day 90) COHb (%)	73	117	46
(Day 90) C Reactive Protein (mg/L)	73	117	46
(Day 90) Diastolic Blood Pressure	73	117	46
(Day 90) Exhaled CO (ppm)	73	117	46
(Day 90) Fibrinogen (mg/dL)	73	117	46
(Day 90) HDL Cholesterol (mg/dL)	73	117	46
(Day 90) HEMA (pg/mg creat)	73	117	46
(Day 90) Hemoglobin A1C (%)	73	117	46
(Day 90) HMPMA (ng/mg creat)	73	117	46
(Day 90) Intercellular Adhesion Molecule 1 (ng/mL)	73	117	46
(Day 90) LDL Cholesterol (mg/dL)	73	117	46
(Day 90) Leukocytes (GI/L)	73	117	46
(Day 90) MHBMA (pg/mg creat)	73	117	46
(Day 90) NNN (pg/mg creat)	73	117	46
(Day 90) o-tol (pg/mg creat)	72	117	46
(Day 90) Percent of predicted FEV1 value	41	70	37
(Day 90) Percent of predicted FVC value	41	70	37
(Day 90) Prostaglandin F2 Alpha (pg/mg creat)	73	117	46
(Day 90) Ratio between FEV1/FVC (Derived)	72	117	46
(Day 90) S-PMA (pg/mg creat)	73	117	46
(Day 90) Systolic Blood Pressure	73	117	46
(Day 90) Total 1-OHP (pg/mg creat)	73	117	46
(Day 90) Total NNAL (pg/mg creat)	73	117	46
(Day 90) Triglycerides (mg/dL)	73	117	46

The basic statistics for each arm are displayed in Table 2.

Table 2: Descriptive statistics for the biomarkers in each arm. Med: median, Q1: first quartile, Q3: third quartile.

	CC/mCC Q1	SA Q1	THS/THSm 2.2 Q1	CC/mCC Med	SA Med	THS/THSm 2.2 Med	CC/mCC Q3	SA Q3	THS/THSm 2.2 Q3
(Day 5) CARBXHGB	-0.13	-1.39	-1.32	0	-1.03	-1	-0.13	-1.39	-1.32
(Day 5) CO	-0.27	-2.4	-2.34	-0.07	-1.9	-1.9	-0.27	-2.4	-2.34
(Day 5) U1NACRE	-0.16	-3.62	-3.5	0.06	-3.15	-3.1	-0.16	-3.62	-3.5
(Day 5) U1OHPCRE	-0.33	-1.24	-1.23	-0.2	-0.99	-0.98	-0.33	-1.24	-1.23
(Day 5) U2NACRE	-0.2	-2.44	-2.5	-0.02	-2.06	-2.07	-0.2	-2.44	-2.5
(Day 5) U3HPMCRE	-0.28	-1.66	-1.08	-0.02	-1.28	-0.8	-0.28	-1.66	-1.08
(Day 5) U4ABPCRE	-0.18	-2.21	-2.25	0.01	-1.79	-1.79	-0.18	-2.21	-2.25
(Day 5) UBAPCRE	-0.27	-2	-1.88	-0.08	-1.57	-1.42	-0.27	-2	-1.88
(Day 5) UCEMACRE	-0.27	-2.19	-2.07	-0.07	-1.91	-1.85	-0.27	-2.19	-2.07
(Day 5) UHEMACRE	-0.41	-1.51	-1.41	-0.16	-1.04	-1.01	-0.41	-1.51	-1.41
(Day 5) UHMPMCRE	-0.35	-1.84	-1.73	-0.14	-1.48	-1.3	-0.35	-1.84	-1.73
(Day 5) UMHBMCRE	-0.34	-3	-2.85	-0.07	-2.2	-2.24	-0.34	-3	-2.85
(Day 5) UNNALCRE	-0.16	-1.15	-1.01	-0.03	-0.99	-0.81	-0.16	-1.15	-1.01
(Day 5) UNNNCRE	-0.24	-4.15	-1.93	0.06	-3.45	-1.4	-0.24	-4.15	-1.93
(Day 5) UOTOLCRE	-0.32	-1.46	-1.36	-0.03	-1.03	-0.98	-0.32	-1.46	-1.36
(Day 5) USPMACRE	-0.2	-2.84	-2.77	0.03	-2.43	-2.39	-0.2	-2.84	-2.77
(Day 90) 11-Dehydro-Thromboxane B2 (pg/mg creat)	-0.22	-0.38	-0.42	-0.08	-0.25	-0.22	-0.22	-0.38	-0.42
(Day 90) 1-NA (pg/mg creat)	-0.17	-3.46	-3.32	-0.02	-2.3	-2.6	-0.17	-3.46	-3.32
(Day 90) 2-NA (pg/mg creat)	-0.2	-2.23	-2.39	-0.05	-1.79	-1.85	-0.2	-2.23	-2.39
(Day 90) 3-HPMA (ng/mg creat)	-0.26	-1.39	-0.96	0.06	-0.94	-0.69	-0.26	-1.39	-0.96
(Day 90) 4-ABP (pg/mg creat)	-0.16	-1.94	-1.87	0.12	-1.2	-1.41	-0.16	-1.94	-1.87
(Day 90) B[a]P (fg/mg creat)	-0.25	-1.45	-1.63	0.07	-0.8	-1.03	-0.25	-1.45	-1.63
(Day 90) CEMA (ng/mg creat)	-0.18	-3.01	-2.8	0.06	-2.34	-2.26	-0.18	-3.01	-2.8
(Day 90) Cholesterol (mg/dL)	-0.08	-0.06	-0.11	-0.03	0.01	-0.02	-0.08	-0.06	-0.11
(Day 90) COHb (%)	-0.12	-0.77	-0.98	0.02	-0.58	-0.68	-0.12	-0.77	-0.98
(Day 90) C Reactive Protein (mg/L)	-0.23	0	-0.36	0	0	0	-0.23	0	-0.36
(Day 90) Diastolic Blood Pressure	-0.11	-0.09	-0.11	-0.06	-0.03	-0.06	-0.11	-0.09	-0.11
(Day 90) Exhaled CO (ppm)	-0.45	-2.31	-2.22	0	-1.75	-1.73	-0.45	-2.31	-2.22
(Day 90) Fibrinogen (mg/dL)	-0.15	-0.08	-0.17	-0.03	-0.01	-0.04	-0.15	-0.08	-0.17
(Day 90) HDL Cholesterol (mg/dL)	-0.09	-0.02	-0.04	-0.02	0.06	0.06	-0.09	-0.02	-0.04
(Day 90) HEMA (pg/mg creat)	-0.12	-1.18	-1.12	0.09	-0.67	-0.7	-0.12	-1.18	-1.12
(Day 90) Hemoglobin A1C (%)	-0.02	-0.04	-0.02	0	0	0	-0.02	-0.04	-0.02
(Day 90) HMPMA (ng/mg creat)	-0.36	-1.26	-1.28	-0.14	-0.66	-0.87	-0.36	-1.26	-1.28
(Day 90) Intercellular Adhesion Molecule 1 (ng/mL)	-0.08	-0.29	-0.21	0.01	-0.12	-0.11	-0.08	-0.29	-0.21
(Day 90) LDL Cholesterol (mg/dL)	-0.16	-0.13	-0.19	-0.06	-0.01	-0.08	-0.16	-0.13	-0.19
(Day 90) Leukocytes (GI/L)	-0.16	-0.29	-0.22	-0.05	-0.1	-0.07	-0.16	-0.29	-0.22
(Day 90) MHBMA (pg/mg creat)	-0.24	-2.24	-2.3	-0.02	-1.37	-1.64	-0.24	-2.24	-2.3
(Day 90) NNN (pg/mg creat)	-0.45	-3.73	-2.3	0.02	-2.82	-1.56	-0.45	-3.73	-2.3
(Day 90) o-tol (pg/mg creat)	-0.38	-1	-1.26	-0.01	-0.74	-0.82	-0.38	-1	-1.26
(Day 90) Percent of predicted FEV1 value	-0.05	-0.01	-0.01	0	0.02	0.02	-0.05	-0.01	-0.01
(Day 90) Percent of predicted FVC value	-0.05	-0.02	-0.02	0	0.03	0.01	-0.05	-0.02	-0.02
(Day 90) Prostaglandin F2 Alpha (pg/mg creat)	-0.03	-0.2	-0.17	0.12	0.06	-0.02	-0.03	-0.2	-0.17
(Day 90) Ratio between FEV1/FVC (Derived)	-0.03	-0.02	-0.02	-0.01	0	0	-0.03	-0.02	-0.02
(Day 90) S-PMA (pg/mg creat)	-0.33	-2.85	-2.53	-0.04	-1.95	-1.95	-0.33	-2.85	-2.53
(Day 90) Systolic Blood Pressure	-0.11	-0.07	-0.1	-0.02	-0.04	-0.04	-0.11	-0.07	-0.1
(Day 90) Total 1-OHP (pg/mg creat)	-0.03	-0.77	-0.87	0.13	-0.47	-0.49	-0.03	-0.77	-0.87
(Day 90) Total NNAL (pg/mg creat)	-0.1	-2.76	-1.87	0.2	-1.46	-1.31	-0.1	-2.76	-1.87
(Day 90) Triglycerides (mg/dL)	-0.1	-0.07	-0.2	0.03	0.02	0.01	-0.1	-0.07	-0.2

4.2 Biomarker Boxplots

Boxplots for the data of the study REX are displayed in Figure 2.

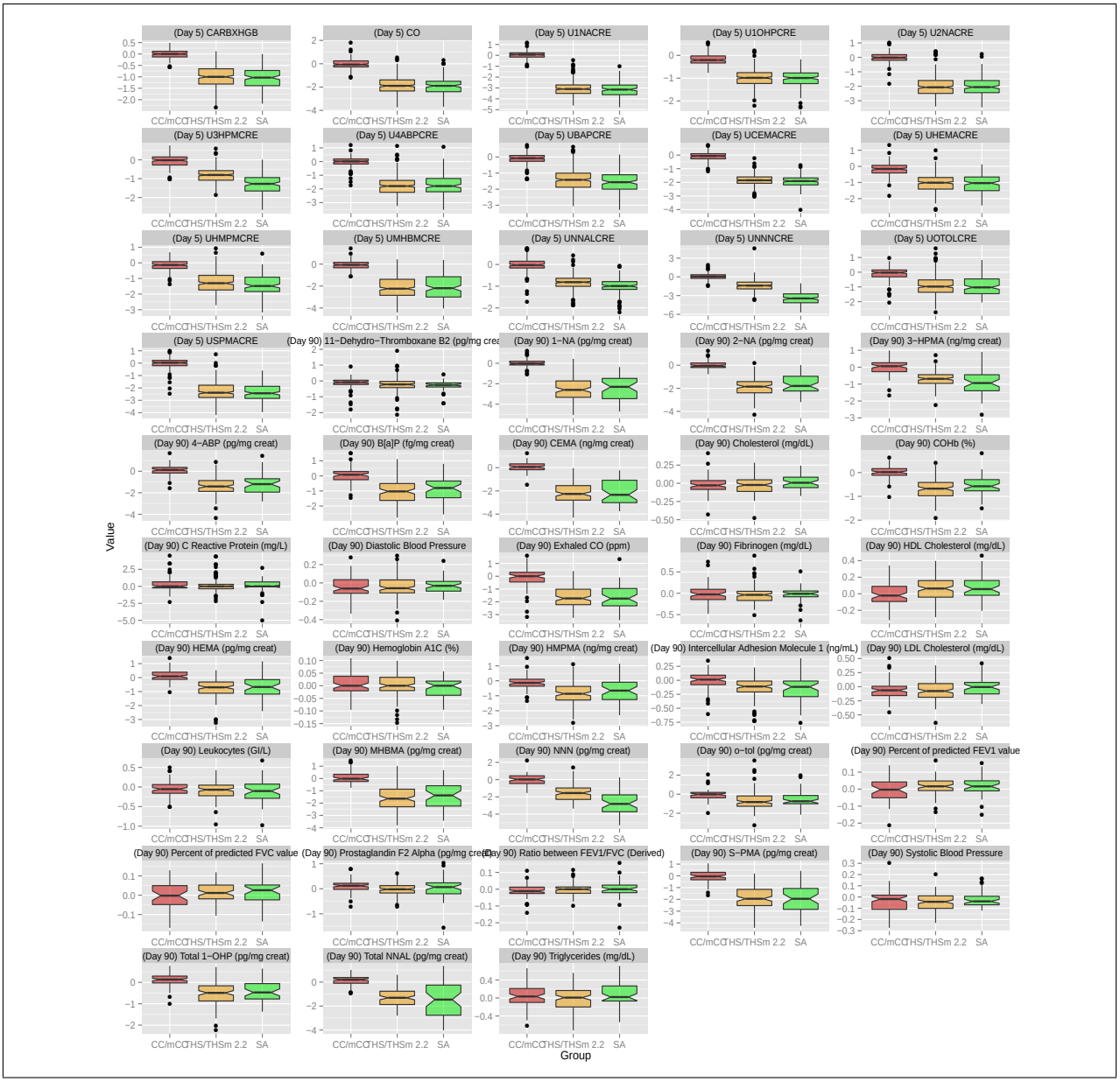


Figure 2: Boxplots the endpoints under consideration

4.3 Checking the Betweenness Assumption

A first check is performed in order to fulfill the assumption 2 ("betweenness"). For that purpose, the posterior distribution of $RD_i|CC - RD_i|Cess$ is sampled and only those markers for which the 5%-quantile is above 0 are considered for the F.

```
## Unknown suffciency:0.5
```

The markers NOT satisfying the assumption 2 are

1. `(Day 90) Cholesterol (mg/dL)`
2. `(Day 90) C Reactive Protein (mg/L)`
3. `(Day 90) Fibrinogen (mg/dL)`

4. `(Day 90) LDL Cholesterol (mg/dL)`
5. `(Day 90) Percent of predicted FEV1 value`
6. `(Day 90) Percent of predicted FVC value`
7. `(Day 90) Prostaglandin F2 Alpha (pg/mg creat)`
8. `(Day 90) Ratio between FEV1/FVC (Derived)`

5 Posterior distribution for RD_i

Analytical posterior distributions were sampled for the group means. Only samples for which the ordering constraint (assumption 2) is satisfied were retained. Indeed, if the range of values for the parameter of interest θ is restricted (i.e $\pi(\theta) > 0 \Leftrightarrow \theta \in \Theta_c$) then by Bayes' theorem, we consider: $\pi(\theta|\boldsymbol{x}, \theta \in \Theta_c) = \begin{cases} \frac{\pi(\theta|\boldsymbol{x})}{\int_{\Theta_c} \pi(\theta|\boldsymbol{x}) d\theta} & \text{if } \theta \in \Theta_c \\ 0 & \text{else} \end{cases}$

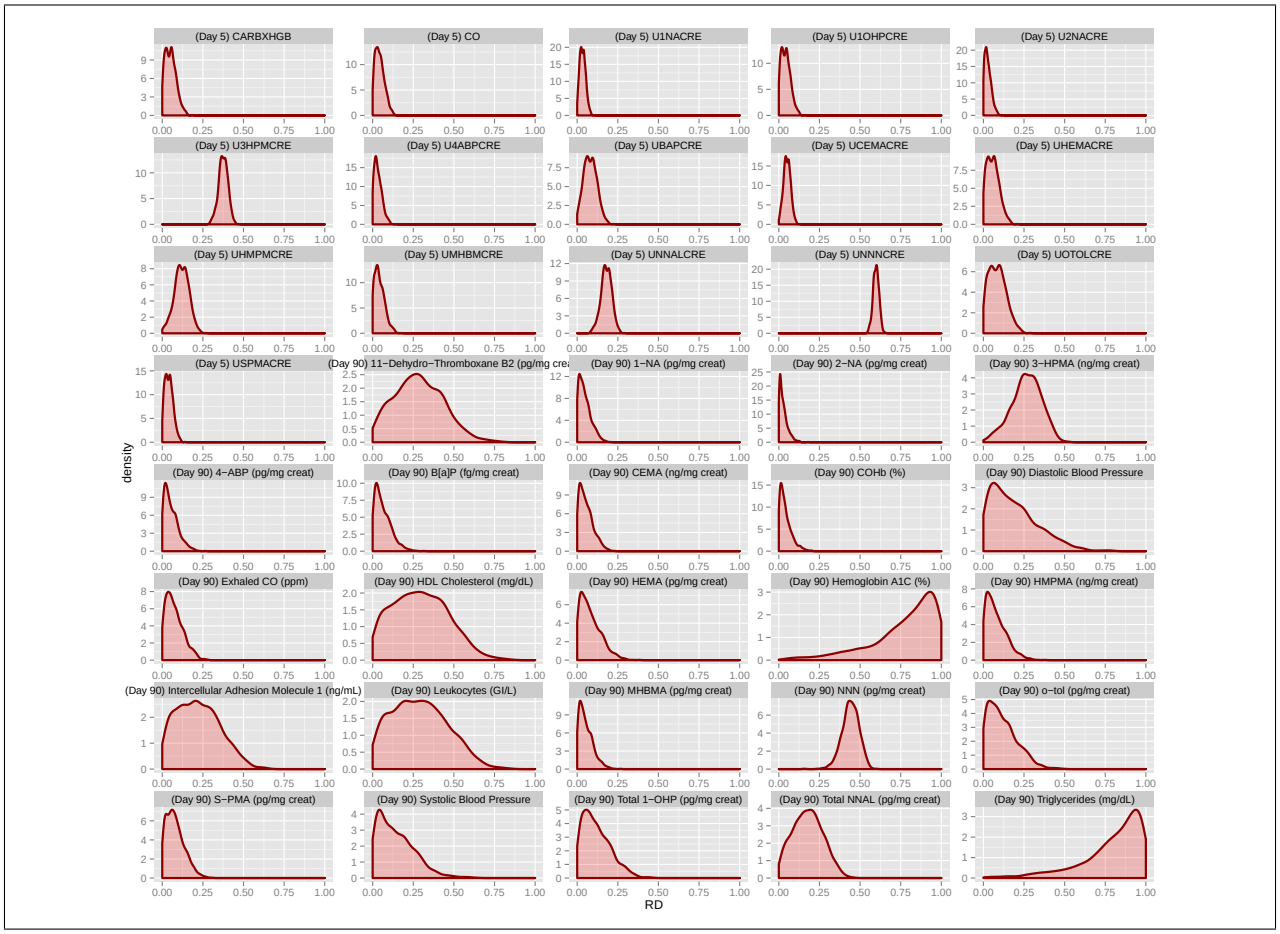


Figure 3: Posterior distributions for the endpoints under consideration.

6 Sampling the link functions

6.1 Sampling the β's

We have defined a family of factors $\psi_i(\beta_i, x_i)$ in the expansion of ψ which depends on the interpretable parameters β_i. In absence of any further quantifiable knowledge in these parameters, we will sample β_i independently uniformly from [0.5,1]. In this study, the necessity was set to 0.5.

6.2 Sampling the main coefficients

The final F is assumed to be an expansion of the form

$$F = \alpha_0 \psi_0(U) + \sum_i \alpha_i \psi_i(RD_i)$$

where $\sum_i \alpha_i = 1$. The ignorance about the parameters α_i will be translated by sampling from the most uniform Dirichlet distribution while keeping the characterization of the sufficiency of the set of markers considered (Table 3). This in particular implies that the parameter α_0 is thought to be such that $\frac{\alpha_0}{\alpha_0 + \sum \alpha_i} = 0.5$.

Table 3: Individual sufficiency values.

	ID	Value
0	U	0.5
1	(Day 5) CARBXHGB	0.01
2	(Day 5) CO	0.01
3	(Day 5) U1NACRE	0.01
4	(Day 5) U1OHPCRE	0.01
5	(Day 5) U2NACRE	0.01
6	(Day 5) U3HPMCRE	0.01
7	(Day 5) U4ABPCRE	0.01
8	(Day 5) UBAPCRE	0.01
9	(Day 5) UCEMACRE	0.01
10	(Day 5) UHEMACRE	0.01
11	(Day 5) UHMPMCRE	0.01
12	(Day 5) UMHBMCRE	0.01
13	(Day 5) UNNALCRE	0.01
14	(Day 5) UNNNCRE	0.01
15	(Day 5) UOTOLCRE	0.01
16	(Day 5) USPMACRE	0.01
17	(Day 90) 11-Dehydro-Thromboxane B2 (pg/mg creat)	0.01
18	(Day 90) 1-NA (pg/mg creat)	0.01
19	(Day 90) 2-NA (pg/mg creat)	0.01
20	(Day 90) 3-HPMA (ng/mg creat)	0.01
21	(Day 90) 4-ABP (pg/mg creat)	0.01
22	(Day 90) B[a]P (fg/mg creat)	0.01
23	(Day 90) CEMA (ng/mg creat)	0.01
24	(Day 90) COHb (%)	0.01
25	(Day 90) Diastolic Blood Pressure	0.01
26	(Day 90) Exhaled CO (ppm)	0.01
27	(Day 90) HDL Cholesterol (mg/dL)	0.01
28	(Day 90) HEMA (pg/mg creat)	0.01
29	(Day 90) Hemoglobin A1C (%)	0.01
30	(Day 90) HMPMA (ng/mg creat)	0.01
31	(Day 90) Intercellular Adhesion Molecule 1 (ng/mL)	0.01
32	(Day 90) Leukocytes (GI/L)	0.01
33	(Day 90) MHBMA (pg/mg creat)	0.01
34	(Day 90) NNN (pg/mg creat)	0.01
35	(Day 90) o-tol (pg/mg creat)	0.01
36	(Day 90) S-PMA (pg/mg creat)	0.01
37	(Day 90) Systolic Blood Pressure	0.01
38	(Day 90) Total 1-OHP (pg/mg creat)	0.01
39	(Day 90) Total NNAL (pg/mg creat)	0.01
40	(Day 90) Triglycerides (mg/dL)	0.01

7 Results

7.1 Final distribution for F

When being uninformed, the initial assumption on F states that it follows an uniform distribution between 0 and 1 ($\mathcal{U}([0,1])$); that is all the values between 0 and 1 are equally plausible for F. Therefore the distribution from the various scenarios considered has to be compared to this initial assumption (represented as a grey box in Figure 5).

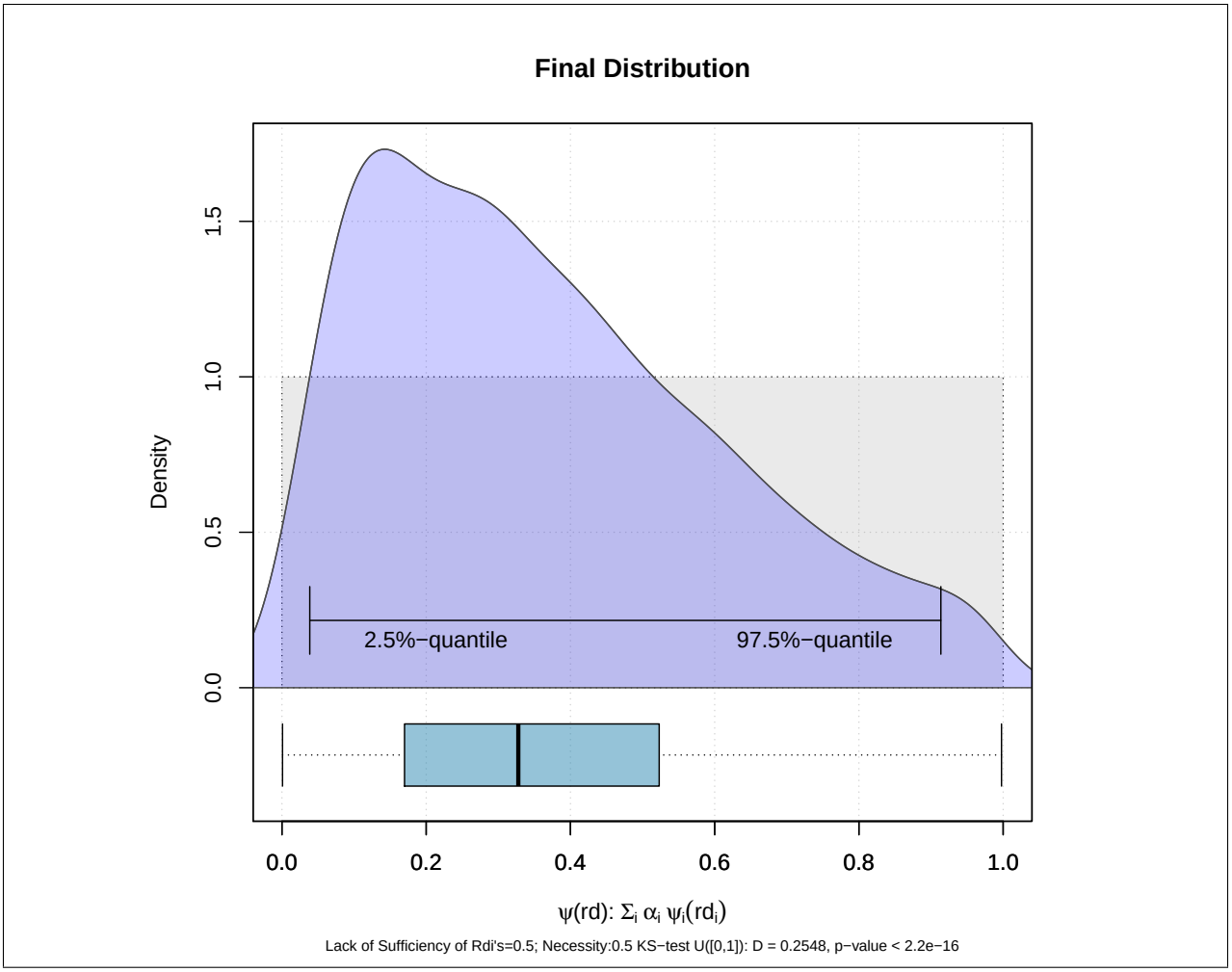


Figure 4: Overall Posterior distribution from the sampled link functions. The grey distribution indicate the uniform distribution which is the null distribution for the F factor $\mathcal{U}(0,1)$. The interval bounds are the 2.5% - 97.5% quantiles. A boxplot is shown as an alternative representation of the distribution.

7.2 Example of application to the PHIM

Some example for the ER model

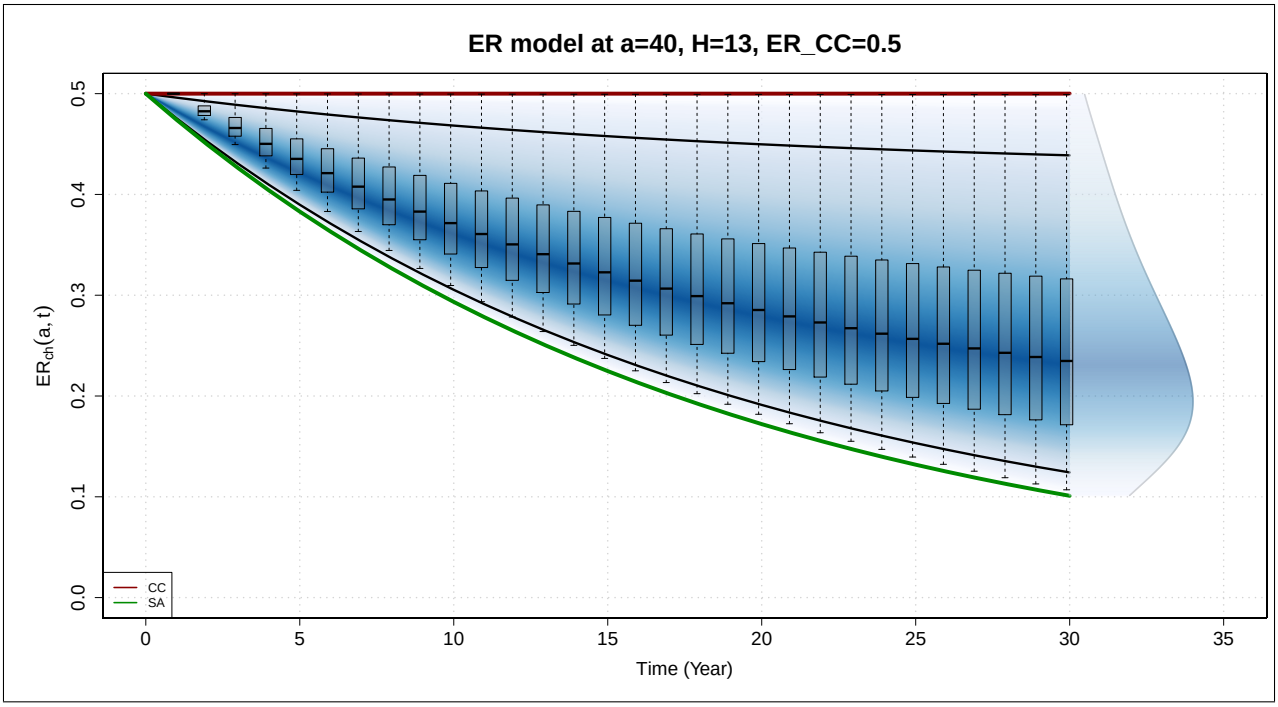


Figure 5: ER model for F sampled from the density estimate. Black lines represents the 5-th and 95-th quantile curves.

8 Annexes

```
## R version 3.1.2 (2014-10-31)
## Platform: x86_64-unknown-linux-gnu (64-bit)
##
## locale:
##  [1] LC_CTYPE=en_US.utf8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.utf8        LC_COLLATE=en_US.utf8
##  [5] LC_MONETARY=en_US.utf8    LC_MESSAGES=en_US.utf8
##  [7] LC_PAPER=en_US.utf8       LC_NAME=C
##  [9] LC_ADDRESS=C              LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.utf8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] tools     stats     graphics  grDevices utils     datasets
## [7] methods   base
##
## other attached packages:
##  [1] plotrix_3.5-11     rctrack_1.2        R.utils_1.34.0
##  [4] R.oo_1.18.0        R.methodsS3_1.6.1  markdown_0.7.4
##  [7] supraHex_1.4.0     hexbin_1.27.0      FfactorPM_0.0.3
## [10] CDBUtils_0.0.8     xlsx_0.5.7         xlsxjars_0.6.1
## [13] rJava_0.9-6        tm_0.6             NLP_0.1-6
## [16] RGraph2js_0.1.4    digest_0.6.8       rjson_0.2.15
## [19] whisker_0.3-2      xtable_1.7-4       RlsfPM_0.99.5
## [22] plyr_1.8.1         sets_1.0-13        knitr_1.9
## [25] ImagePlotPM_0.0.5  gplots_2.14.2      RColorBrewer_1.0-5
## [28] reshape2_1.4.1     MCMCpack_1.3-3     MASS_7.3-35
## [31] coda_0.16-1        lattice_0.20-29    ggplot2_1.0.0
## [34] colorout_1.1-0
##
## loaded via a namespace (and not attached):
##  [1] ape_3.1-4          bitops_1.0-6       caTools_1.17.1
```

```
##  [4] colorspace_1.2-4   evaluate_0.5.5     formatR_1.0
##  [7] gdata_2.13.3       grid_3.1.2         gtable_0.1.2
## [10] gtools_3.4.1       highr_0.4          KernSmooth_2.23-13
## [13] labeling_0.3       magrittr_1.5       munsell_0.4.2
## [16] nlme_3.1-118       parallel_3.1.2     proto_0.3-10
## [19] Rcpp_0.11.3        scales_0.2.4       slam_0.1-32
## [22] stringi_1.0-1      stringr_1.0.0
```